# Kernel Regression Estimation Using Repeated Measurements Data

JEFFREY D. HART and THOMAS E. WEHRLY*

The estimation of growth curves has been studied extensively in parametric situations. Here we consider the nonparametric estimation of an average growth curve. Suppose that there are observations from several experimental units, each following the regression model $y(x_j) = f(x_j) + \varepsilon_j$ $(j = 1, \ldots, n)$, where $\varepsilon_1, \ldots, \varepsilon_n$ are correlated zero mean errors and $0 \le x_1 < \cdots < x_n \le 1$ are fixed constants. We study some of the properties of a kernel estimator of $f(x)$. Asymptotic and finite-sample results concerning the mean squared error of the estimator are obtained. In particular, the influence of correlation on the bandwidth minimizing mean squared error is discussed. A data-based method for selecting the bandwidth is illustrated in a data analysis.

Most previous research on kernel regression estimators has involved uncorrelated errors. We investigate how dependence of the errors changes the behavior of a kernel estimator. Our theorems concerning the asymptotic mean squared error show that the estimator cannot be consistent unless the number of experimental units tends to infinity. This contrasts with the results for uncorrelated errors, where the estimator is consistent when the number of distinct $x$ values tends to infinity. Finite-sample results indicate that the optimum bandwidth when the errors are correlated can be either larger or smaller than the optimum bandwidth with uncorrelated errors. If the number of $x$ values is not large and/or the errors are highly positively correlated, the optimum bandwidth tends to be smaller than when the errors are uncorrelated. This is contrary to existing examples wherein serially correlated errors require larger than usual bandwidths.

In our data analysis, we choose the bandwidth that minimizes an estimate of the mean average squared error while taking into account the presence of correlated errors. Using the same data we show that ignoring correlation leads to an oversmoothed kernel estimate. An analytic result illustrates that this phenomenon is not necessarily an anomaly of the data.

KEY WORDS: Nonparametric regression; Growth curves; Correlated data; Optimum bandwidth.

## 1. INTRODUCTION

Suppose that $f$ is a regression function defined on the interval [0, 1]. A model often considered in the study of nonparametric estimators of $f$ is

$$y_i = f(x_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n, \tag{1.1}$$

where the $y_i$'s are observed random variables, the $x_i$'s are known constants with $0 \le x_1 < x_2 < \cdots < x_n \le 1$, and the $\varepsilon_i$'s are independent random variables with mean 0 and variance $\sigma^2$. Of much interest recently have been kernel estimators of $f$ such as that proposed by Gasser and Müller (1979), namely

$$\hat{f}_h(x) = h^{-1} \sum_{i=1}^{n} y_i \int_{x_{i-1}}^{x_i} K((x - u)/h)\, du, \tag{1.2}$$

where $K$ is usually a density function symmetric about 0 and the bandwidth $h$ is positive. For reviews of properties of $\hat{f}_h$ and other kernel-type estimators, see Gasser and Müller (1984) and Prakasa Rao (1983). In addition, the reader is referred to Rice (1984a) and Härdle and Marron (1985) for results on the important problem of choosing the bandwidth from the observed data.

In the current article, we consider a nonparametric approach to problems in which a random sample of $m$ experimental units is available and the observed data for the $i$th unit are the values, $y_i(x_j)$ $(j = 1, \ldots, n)$, of a response variable corresponding to the values, $x_j$ $(j = 1, \ldots, n)$, of a controlled variable. The function to be estimated is $f$, where $f(x)$ is the population mean response corresponding to $x \in [0, 1]$. As indicated by Grizzle and Allen (1969), such situations arise naturally in the analysis of growth and dose response curves. For example, one may wish to estimate average organism size (length, weight, etc.) as a function of age for some population of organisms.

It is clear that the observations $y_i(x_j)$ $(j = 1, \ldots, n)$ made on the same experimental unit will in general be correlated. Hence it is of interest to relax the assumption of independent errors in a model such as (1.1) and to investigate the resulting properties of kernel estimators of the mean response function $f$. Such investigations have been conducted for parametric situations by, among others, Potthoff and Roy (1964), Rao (1965), Grizzle and Allen (1969), Geisser (1980), and Reinsel (1982). Ghosh, Grizzle, and Sen (1973) studied nonparametric methodology in longitudinal studies. Härdle and Tuan (1986) investigated using $M$-estimate-type smoothers on time series data. In addition, Gasser, Müller, Kohler, Molinari, and Prader (1984) considered the use of kernel estimators in the analysis of growth curves. Their work, however, applies to the estimation of an individual's growth curve, and thus the problem of correlated errors does not arise.

The remainder of the article will proceed as follows. In Section 2, a model, analogous to (1.1) but allowing for correlation among the errors, is proposed. In Section 3, we discuss our analysis of a data set considered by Andersen, Jensen, and Schou (1981) and Azzalini (1984). In Section 4, asymptotic mean squared error (MSE) properties of a kernel estimator under the correlation model will be considered. In particular, results are obtained concerning the optimum choice of a bandwidth. Section 5 contains numerical results that show that choosing the bandwidth of the kernel estimator as if observations were uncorrelated can be very inefficient relative to using the optimum bandwidth. It is also demonstrated that, depending on the sizes of $n$ and the correlation, the optimum bandwidth with correlated errors can be either larger or smaller than with

* Jeffrey D. Hart is Assistant Professor and Thomas E. Wehrly is Associate Professor, Department of Statistics, Texas A&M University, College Station, TX 77843. This work was supported in part by Office of Naval Research Contract N00014-85-K-0723. The authors thank S. Cambanis for his helpful insight concerning the consistency of the kernel estimator. They are also grateful to the referees and an associate editor, whose comments led to an improvement of this article.


Journal of the American Statistical Association
December 1986, Vol. 81, No. 396, Theory and Methods

uncorrelated errors. Finally, in Section 6 we discuss some areas for future research.

## 2. THE MODEL TO BE CONSIDERED

The following model will be considered throughout the rest of our discussion. Similar models have appeared in the literature, for example, Azzalini's (1984) model for analyzing collections of time series. Define

$$y_i(x_j) = f(x_j) + \varepsilon_i(x_j), \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \quad (2.1)$$

where the $x_j$'s are fixed with $0 \le x_1 < x_2 < \cdots < x_n \le 1$ and the $\varepsilon_i(x_j)$'s are zero mean random variables satisfying

$$\begin{aligned} \operatorname{cov}(\varepsilon_i(x_j), \varepsilon_k(x_l)) &= \sigma^2 \rho(x_j - x_l), & i = k \\ &= 0, & i \neq k. \quad (2.2) \end{aligned}$$

The correlation function $\rho$ is even with $\rho(0) = 1$ and $|\rho(y)| \le 1$ for all $y \in [-1, 1]$. For our theoretical results, additional assumptions will be required concerning the smoothness of $\rho$.

If $y_i(x_1), \ldots, y_i(x_n)$ are observations made on the $i$th of $m$ organisms, then this model implies that observations made on the same organism are correlated and those made on different organisms are uncorrelated. Such a model certainly seems reasonable if the $m$ organisms are a random sample from some population of organisms. We also point out that our model could be generalized in more than one way. For example, it may be that the available experimental units are not all measured at the same $x$ values. In addition, the correlations between observations made on the same unit need not satisfy the stationarity assumption in (2.2). For the sake of clarity, however, we prefer to restrict attention in this article to the simpler model defined by (2.1) and (2.2).

Before continuing, we make the following observation. Defining $\bar{y}(x_j) = \sum_{i=1}^m y_i(x_j)/m$, we have [from (2.1) and (2.2)]

$$\bar{y}(x_j) = f(x_j) + \bar{\varepsilon}(x_j), \quad (2.3)$$

where

$$\bar{\varepsilon}(x_j) = \sum_{i=1}^m \varepsilon_i(x_j)/m,$$

$$\operatorname{cov}(\bar{\varepsilon}(x_j), \bar{\varepsilon}(x_k)) = (\sigma^2/m)\rho(x_j - x_k). \quad (2.4)$$

Expressing the model in this way is useful, since the problem of estimating $f$ may now be regarded as that of fitting a smooth curve through the sample means $\bar{y}(x_j)$. The estimator we will investigate is

$$\hat{f}_h(x) = h^{-1} \sum_{j=1}^n \bar{y}(x_j) \int_{s_{j-1}}^{s_j} K((x - u)/h)\, du, \quad (2.5)$$

where $s_0 = 0$, $s_j = (x_j + x_{j+1})/2$ $(j = 1, \ldots, n - 1)$, $s_n = 1$, and $K$ is a density function with support $[-1, 1]$.

## 3. DATA ANALYSES

As motivation for the rest of the article, we here present our analysis of a data set considered by Andersen et al. (1981). The data are plasma citrate concentrations measured the same day on each of $m = 10$ human subjects. The measurements on an individual were taken each hour from 8:00 a.m. to 9:00 p.m., so $n = 14$. Our interest is in using these data to estimate the population mean plasma citrate concentration as a function of time of day. Although the number of design points is somewhat small, we feel that this example illustrates well a method for adjusting a kernel estimate to account for correlated observations.

The estimate to be employed is a slightly modified version of (2.5) that takes into account boundary effects. Defining

$$a_h(x) = h^{-1} \sum_{j=1}^n \int_{s_{j-1}}^{s_j} K\left(\frac{x - u}{h}\right) du,$$

the modified estimate is

$$\hat{f}_h^*(x) = \hat{f}_h(x)/a_h(x), \quad (3.1)$$

which is equal to $\hat{f}_h(x)$ except when $x$ is within a bandwidth (i.e., $h$) of either endpoint. Using (3.1) ensures that one's estimate is always a weighted average of $\bar{y}_i$'s, thus reducing bias near the boundary. For more refined boundary modifications, see Gasser and Müller (1979) or Rice (1984b).

To employ an estimate of the form (3.1), one must first select the value of $h$. A method investigated by Rice (1984a) for choosing the bandwidth proceeds by determining the value of $h$ that minimizes an estimated mean average squared error (MASE) curve, where MASE is

$$M(h) = \frac{1}{n} \sum_{j=1}^n E(\hat{f}_h^*(x_j) - f(x_j))^2. \quad (3.2)$$

When model (2.1)–(2.2) holds, it may be shown that

$$M(h) = 1/n[E(RSS(h))] - \sigma^2/m[1 - 2/n \operatorname{tr}(K_h R)],$$

where

$$RSS(h) = \sum_{j=1}^n (\bar{y}(x_j) - \hat{f}_h^*(x_j))^2,$$

$K_h$ is an $n \times n$ matrix with $(i, j)$th element

$$h^{-1} \int_{s_{j-1}}^{s_j} K\left(\frac{x_i - u}{h}\right) du/a_h(x_i),$$

and $R$ is a matrix of correlations with $(i, j)$th element $\rho(x_i - x_j)$. For equally spaced $x$'s and a symmetric (about 0) kernel, we have, defining $\Delta = x_2 - x_1$,

$$\operatorname{tr}(K_h R) = 2b_{n,h}(0) \int_0^{\Delta/(2h)} K(y)\, dy + 2 \sum_{i=1}^{n-1} b_{n,h}(i)\rho(i\Delta) \int_{(i-1/2)\Delta/h}^{(i+1/2)\Delta/h} K(y)\, dy, \quad (3.3)$$

where $b_{n,h}(i) = \sum_{j=1}^{n-i} 1/a_h(x_j)$. If one uses a kernel with support $[-1, 1]$, it follows that to estimate $\operatorname{tr}(K_h R)$ for $h \le h_{\max}$ one need only estimate $\rho(k\Delta)$ for $k = 1, \ldots, k_{\max}$ = largest integer less than $h_{\max}/\Delta + \frac{1}{2}$.

For the plasma citrate data, $\rho(k)$ ($\Delta = 1$ hour) was estimated by

$$\hat{\rho}(k) = \hat{c}(k)/\hat{c}(0), \qquad k = 1, \ldots, 7,$$

where

$$\hat{c}(k) = (nm)^{-1} \sum_{i=1}^{m} \sum_{j=1}^{n-k} (y_i(x_j) - \bar{y}(x_j)) \times (y_i(x_{j+k}) - \bar{y}(x_{j+k})), \qquad k = 0, \ldots, 7.$$

The estimated autocorrelations are given as follows:

| $k$: | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| $\hat{\rho}(k)$: | .677 | .566 | .488 | .419 | .357 | .333 | .288 |

[When the $x$'s are unequally spaced, methodology used in irregularly spaced time series analysis (see, e.g., Masry 1983) could be used to estimate the correlation function.] Using the Epanechnikov kernel $K(y) = .75(1 - y^2)I_{(-1,1)}(y)$, the MASE curve was estimated by

$$\hat{M}(h) = RSS(h)/n - (\hat{c}(0)/m)[1 - (2/n)\hat{\text{tr}}(K_h R)],$$

where $\hat{\text{tr}}(K_h R)$ is simply (3.3) with $\rho(i)$ replaced by $\hat{\rho}(i)$. The curve $\sqrt{\hat{M}}$ is shown in Figure 1 along with $\sqrt{\hat{M}^*}$, where $\hat{M}^*$ is obtained on the assumption that the $\bar{y}(x_j)$'s are uncorrelated; that is,

$$\hat{M}^*(h) = RSS(h)/n - (\hat{c}(0)/m)[1 - 4b_{n,h}(0) \int_0^{\Delta/(2h)} K(y)\, dy/n]. \quad (3.4)$$

The minimum of $\hat{M}$ occurred at $h = 1.11$, and the minimum of $\hat{M}^*$ occurred at $h = 2.03$. The corresponding kernel estimates, $\hat{f}_h^*$, are given in Figure 2.

Contrary to the results of our data analysis, examples of Diggle (1985) and Diggle and Hutchinson (1985) indicate that ignoring correlation tends to give drastically undersmoothed estimates. To explain this seeming contradiction, we first note that positive serial correlation makes the

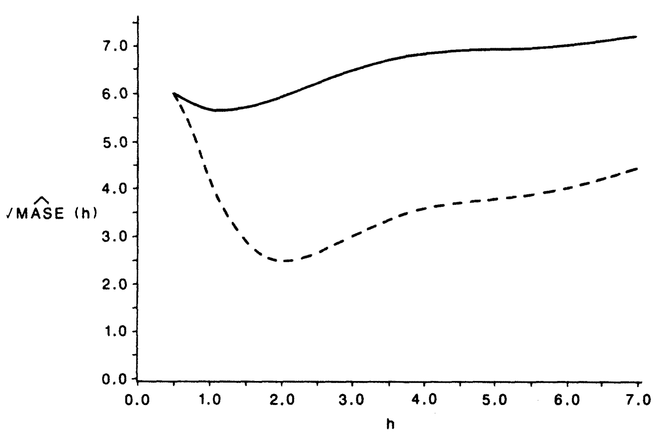


*Figure 1. Two Estimated $\sqrt{MASE}$ Curves for the Plasma Citrate Data. The dashed curve was obtained on the assumption of uncorrelated observations, and the other curve allows for correlation between observations at different design points.*

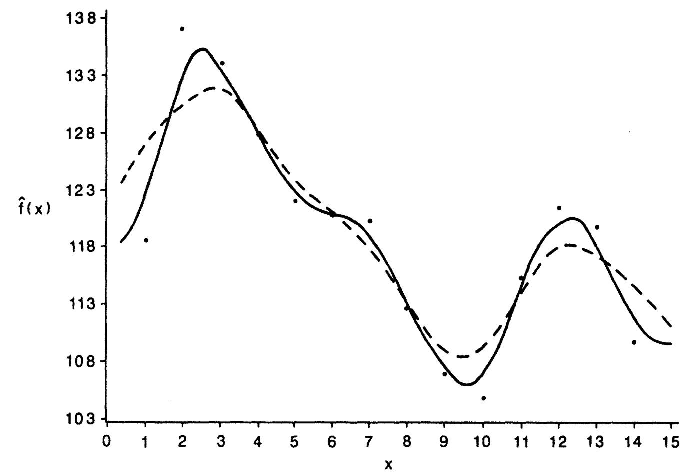


*Figure 2. Regression Function Estimates for the Plasma Citrate Data. The dots denote observed mean plasma citrate concentrations. The dashed curve is a kernel estimate using the estimated optimum bandwidth assuming uncorrelated observations. The other estimate uses an estimated optimum bandwidth allowing for correlation.*

sample path $\{\bar{y}(x_j): j = 1, \ldots, n\}$ smoother than when the errors are uncorrelated. A data-based smoother that assumes uncorrelated errors attributes all of the smoothness in the data to the unknown function and none to correlation. The result is an estimate that very faithfully tracks the data. When $n$ is large, excursions of the error process from 0 would cause such an estimate to be much less smooth than the function $f$. Now suppose that $n$ is somewhat small and that one uses an estimate that essentially interpolates the data. This sort of estimate would be much too wiggly if the errors are uncorrelated. If the errors are sufficiently highly correlated, however, an interpolating estimate would be quite smooth relative to the unknown function. This intuitive argument suggests that for $n$ small enough and $\rho$ large enough, the optimum bandwidth may actually be smaller than when the errors are uncorrelated. On the other hand, when $n$ is large enough a larger than usual bandwidth will be required. These conjectures will be substantiated in Sections 4 and 5.

To make the previous discussion more concrete, the reader is referred to Figures 3 and 4. In each case, data were simulated from the model $y_j = 10(1 + j/(n + 1)) + \varepsilon_j$ ($j = 1, \ldots, n$), where $\varepsilon_j = .75\varepsilon_{j-1} + z_j$ and the $z_j$'s are iid $N(0, 1)$. Kernel estimates were chosen using the criterion (3.4) (i.e., the correlation was ignored). Figure 3 with $n = 100$ reproduces the result obtained by Diggle (1985). Ignoring correlation produced a much undersmoothed estimate. In Figure 4 with $n = 10$, however, it is seen that the smoothness in the data induced by serial correlation essentially prevents any chance of undersmoothing.

## 4. ASYMPTOTIC MEAN SQUARED ERROR OF $\hat{f}_h(x)$

The statistical literature contains an abundance of results on the asymptotic behavior of kernel regression estimators. Virtually all of this literature, though, deals with the un-

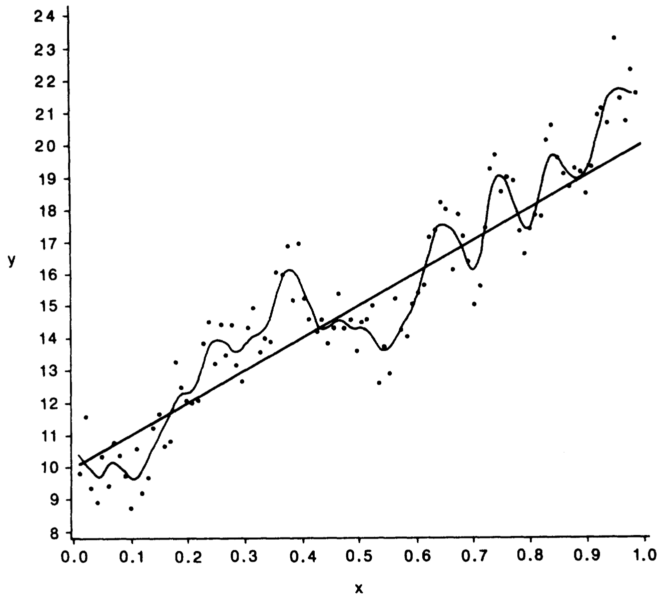


*Figure 3. A Kernel Estimate Chosen Ignoring Serial Correlation, With $n = 100$. The data were generated from the model $y_j = 10(1 + j/(n + 1)) + \varepsilon_j$ ($j = 1, \ldots, n$), where $\varepsilon_j = .75\varepsilon_{j-1} + Z_j$ and the $Z_j$'s are iid $N(0, 1)$. The data are plotted with the line $y = 10(1 + x)$ and a kernel estimate of the form (3.1), where $K$ = Epanechnikov kernel and $h = .027$. The bandwidth of the estimate was chosen using criterion (3.4), which ignores serial correlation.*

correlated errors case. To our knowledge, in the fixed $x$'s setting the only existing work on kernel estimators that takes into account correlated errors is that of Müller (1984) and Härdle and Tuan (1986). In Müller's treatment, however, $\operatorname{corr}(\bar{\varepsilon}(x), \bar{\varepsilon}(x^*))$ $(x \neq x^*)$ tends to 0 as $n \to \infty$, which is an unrealistic assumption in, for example, the growth curve problem. Härdle and Tuan's results require either a slowly varying regression function or an assumption like Müller's. It thus seems of some importance to consider asymptotic properties of a kernel estimator under a model such as our (2.1)–(2.2).

We shall now study the behavior of

$$\operatorname{MSE}(\hat{f}_h(x)) = E(\hat{f}_h(x) - f(x))^2$$

as $n$ and $m \to \infty$ and $h \to 0$. The estimator $\hat{f}_h$ is defined in (2.5), and it will be assumed that $K$ satisfies the Lipschitz condition

$$|K(x) - K(y)| \leq A|x - y| \quad \text{for all } x \text{ and } y \tag{4.1}$$

and some constant $A$. To avoid the trivial estimator $\hat{f}_h(x) \equiv 0$, we insist that $h \geq \min_{1 \leq i \leq n} |x - x_i|$. After stating our results, we make several remarks concerning their statistical relevance. Proofs of the results are given in the Appendix. The proofs of Theorems 1 and 2 use methods like those of Gasser and Müller (1984). Arguing, however, that the bandwidths in Theorems 3 and 4 are asymptotically optimum is somewhat more intricate than in the uncorrelated errors case. This is because $\operatorname{MSE}(\hat{f}_h(x))$ is asymptotic to $\sigma^2/m$ in the cases considered, and thus the second-order efficiency $\operatorname{MSE}(\hat{f}_h(x)) - \sigma^2/m$ must be examined.

*Theorem 1.* Let $0 < x < 1$, and in addition to the previous assumptions on $K$, assume that $\rho$ is Lipschitz continuous in the sense of (4.1) and

$$\max_j |x_j - x_{j-1}| = 0(1/n).$$

We then have, as $n, m \to \infty$ and $h \to 0$,

$$\operatorname{var}(\hat{f}_h(x)) = (\sigma^2/m) \int_{-1}^{1} \int_{-1}^{1} \rho(h(u - v)) \times K(u)K(v)\, du\, dv + 0(1/nm).$$

*Proof.* See the Appendix.

*Theorem 2.* In addition to the assumptions in Theorem 1, assume that $f$ is Lipschitz continuous. We then have, as $n, m \to \infty$ and $h \to 0$,

$$\operatorname{MSE}(\hat{f}_h(x)) = (\sigma^2/m) \int_{-1}^{1} \int_{-1}^{1} \rho(h(u - v)) \times K(u)K(v)\, du\, dv + \left[\int_{-1}^{1} K(u)f(x - hu)\, du - f(x)\right]^2 + o(1/n). \tag{4.2}$$

*Proof.* The result is immediate from our Theorem 1 and equation (6) of Gasser and Müller (1984).

*Theorem 3.* Suppose that $m/n = 0(1)$ as $n, m \to \infty$. Assume also that the conditions of Theorem 1 hold and $f$ is twice continuously differentiable on [0, 1] with $f''(x) \neq$

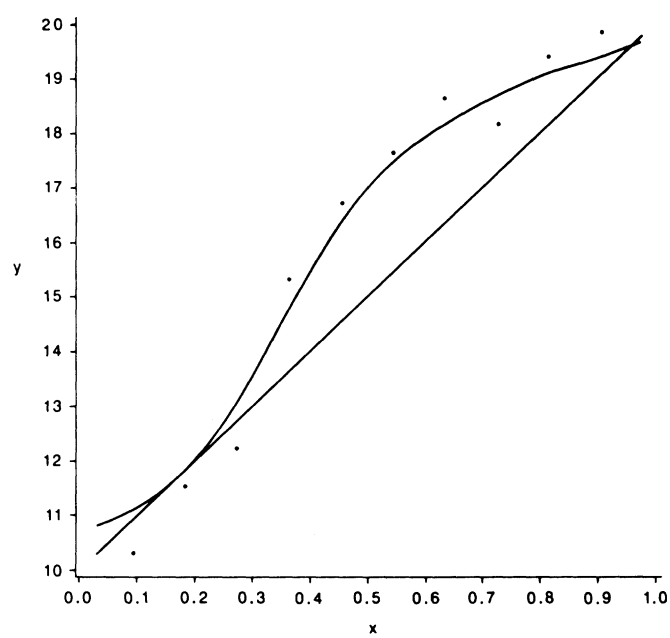


*Figure 4. A Kernel Estimate Chosen Ignoring Serial Correlation, With $n = 10$. The data were generated from the model $y_j = 10(1 + j/(n + 1)) + \varepsilon_j$ ($j = 1, \ldots, n$), where $\varepsilon_j = .75\varepsilon_{j-1} + Z_j$ and the $Z_j$'s are iid $N(0, 1)$. The data are plotted with the line $y = 10(1 + x)$ and a kernel estimate of the form (3.1), where $K$ = Epanechnikov kernel and $h = .227$. The bandwidth of the estimate was chosen using criterion (3.4), which ignores serial correlation.*

0. Now, if $\rho$ has left and right derivatives at 0 with $\rho'(0-) \neq \rho'(0+)$, then as $n, m \to \infty$ and $h \to 0$

$$\mathrm{MSE}(\hat{f}_h(x)) \sim (\sigma^2/m)(1 + \rho'(0+)C_K h) + h^4\sigma_K^4(f''(x))^2/4,$$

where

$$C_K = 2\int_{-1}^{1}\int_{v}^{1}(u - v)K(u)K(v)\,du\,dv$$

and

$$\sigma_K^2 = \int_{-1}^{1} u^2 K(u)\,du.$$

Furthermore, the bandwidth

$$h_m = \left[\frac{\sigma^2\rho'(0-)C_K}{\sigma_K^4(f''(x))^2}\right]^{1/3} m^{-1/3}$$

is optimum in the sense that, for all $n$ and $m$ sufficiently large,

$$\mathrm{MSE}(\hat{f}_{h_m}(x)) < \mathrm{MSE}(\hat{f}_{h_{n,m}}(x)),$$

where $h_{n,m}$ is any sequence of bandwidths tending to 0 as $n, m \to \infty$ and such that $\liminf_{n,m\to\infty}|h_m/h_{n,m} - 1| > 0$.

*Proof.* See the Appendix.

*Theorem 4.* Let all of the conditions of Theorem 3 hold with the exception that $\rho$ is assumed to be twice continuously differentiable on $[-1, 1]$ with $\rho''(0) \neq 0$. Then as $n$, $m \to \infty$ and $h \to 0$

$$\mathrm{MSE}(\hat{f}_h(x)) \sim (\sigma^2/m)(1 + \rho''(0)\sigma_K^2 h^2) + h^4\sigma_K^4(f''(x))^2/4.$$

Furthermore, if in addition $m/n = o(1)$, then the bandwidth

$$h_m^* = \left[\frac{-2\sigma^2\rho''(0)}{\sigma_K^2(f''(x))^2}\right]^{1/2} m^{-1/2}$$

is optimum in the sense of Theorem 3.

*Proof.* The proof proceeds in a manner analogous to that of Theorem 3 and is thus omitted.

*Remarks.*

1. Theorem 2 shows that, with correlated errors, a kernel estimator is not consistent for $f(x)$ unless the number of observations, $m$, at each $x_j$ tends to infinity. With $m$ fixed, if $n \to \infty$, $h \to 0$, and $nh \to \infty$ (the usual prescription for consistency of a kernel estimator), then $\mathrm{MSE}(\hat{f}_h(x)) \to \sigma^2/m$. Of course, in light of the practical problem of interest here, this only makes sense. For a fixed set of $m$ experimental units (animals, say) letting $n \to \infty$ only gives more nearly complete information about the $m$ animals but not about the entire population of animals.

2. As long as $m \to \infty$, Theorem 2 indicates that the condition $nh \to \infty$ is not needed to obtain consistency. Theorems 3 and 4, however, show that when $m$ is not large relative to $n$, $h$ must satisfy $nh \to \infty$ to minimize asymptotically the second-order efficiency $\mathrm{MSE}(\hat{f}_h(x)) - \sigma^2/m$.

3. Comparing $h_m$ and $h_m^*$ in Theorems 3 and 4, it is seen that the less smooth the correlation function $\rho$ is at 0, the larger the optimum bandwidth is. This is intuitively plausible, since when $\rho$ decays quickly to 0, observations somewhat removed from $x$ will contain nonredundant information.

4. Earlier it was claimed that small $n$ and large serial correlation could result in a smaller optimum bandwidth than with uncorrelated errors. If the conditions of Theorem 3 hold but errors are uncorrelated, Gasser and Müller (1984) showed that

$$h_{n,m} = \left[\frac{\sigma^2\int_{-1}^{1}K^2(u)\,du}{\sigma_K^4(f''(x))^2}\right]^{1/5}(nm)^{-1/5}$$

asymptotically minimizes $\mathrm{MSE}(\hat{f}_h(x))$. Comparing this with, say, $h_m$ in Theorem 3, we have $h_m < h_{n,m}$ iff

$$n < b_K\left[\frac{(f''(x))^2}{(\sigma^2/m)(\rho'(0-))^{5/2}}\right]^{2/3}, \tag{4.3}$$

where $b_K$ is a constant depending only on the kernel $K$. Since in Theorem 3 $m/n = 0(1)$, (4.3) fails to hold for all $n$ and $m$ sufficiently large. This is consistent with Diggle's contention that correlated errors require a larger bandwidth. Inequality (4.3) suggests, however, that for $n$ small and serial correlation large [i.e., $\rho'(0-)$ small], the optimum bandwidth may be smaller than usual. The results in the next section prove the last statement conclusively. Inequality (4.3) also indicates that for fixed $\rho$, a lack of smoothness at $x$ [i.e., $|f''(x)|$ large] and/or a small variance $\sigma^2/m$ would tend to yield a smaller bandwidth than with uncorrelated errors.

We have not addressed the situation in which $m/n \to \infty$ as $n, m \to \infty$. Although we will not treat this case fully, it can be verified that if $m/n^2 = o(1)$ and the conditions of either Theorem 3 or 4 hold, then

$$\lim_{\substack{n,m\to\infty\\ h\to 0}} m\,\mathrm{MSE}(\hat{f}_h(x))/\sigma^2 = 1$$

for any sequence of bandwidths satisfying $h = o(m^{-1/4})$. Hence, even in certain cases in which $n$ is small relative to $m$ (e.g., $n = cm^{5/8}$), it is still possible to obtain a kernel estimator that has the same asymptotic MSE as the mean of $m$ observations at the value $x$. It also seems safe to say that the bandwidths $h_m$ and $h_m^*$ in Theorems 3 and 4 are, asymptotically, not small enough to be optimum for the case in which $m/n \to \infty$. This situation is not very interesting from the kernel estimation perspective, since the variance of a mean $\bar{y}(x_j)$ is so small that an estimate interpolating the $\bar{y}$'s would undoubtedly be good asymptotically.

## 5. NUMERICAL RESULTS

A numerical study was done to determine the effect of correlated errors on the optimum bandwidth and the MSE of the estimator. In the case in which the errors are independent, the bandwidth goes to 0 as the number, $n$, of values of the independent variable goes to infinity. Likewise, the MSE goes to 0 and the regression estimator is a consistent estimator of $f(x)$. In the case of correlated errors, however, the limiting MSE (for fixed $h$) can be written as

$$C(h) = \lim_{n\to\infty} \mathrm{MSE}(\hat{f}_h(x)) = \frac{\sigma^2}{m} h^{-2} \int_0^1 \int_0^1 \rho(u - v) \times K\left(\frac{x-u}{h}\right) K\left(\frac{x-v}{h}\right) du\, dv + \left(h^{-1} \int_0^1 f(u) K\left(\frac{x-u}{h}\right) du - f(x)\right)^2 \quad (5.1)$$

using (4.2) of Section 4. We observed that $\lim_{h\to 0} \lim_{n\to\infty} \mathrm{MSE}(\hat{f}_h(x)) = \sigma^2/m$. Thus for a fixed number $m$ of experimental units the MSE does not go to 0 as $h \to 0$. It is natural to ask whether the MSE is minimized for some bandwidth $h > 0$. In this study, we computed the MSE for a grid of bandwidths and found the bandwidth that minimized the MSE. The resulting MSE was also compared with $\sigma^2/m$, the MSE obtained from using a bandwidth of 0 to estimate the function. This is equivalent to using the minimum variance unbiased estimator $\bar{y}(x)$ to estimate $f(x)$. It will be seen that using a biased estimator corresponding to a bandwidth $h > 0$ results in an estimator with smaller MSE than the best unbiased estimator. We wish to emphasize that letting $n \to \infty$ corresponds to having a continuous data record $\{\bar{y}(x): 0 \le x \le 1\}$.

The details of the numerical study follow. The kernel used in the study was the Epanechnikov kernel given by $K(u) = .75(1 - u^2)I_{[-1,1]}(u)$. The two functions in the study were

$$f(x) = 10x^3 - 15x^4 + 6x^5, \qquad 0 < x < 1, \quad (5.2)$$

and

$$f(x) = (x - \tfrac{1}{2})^2, \qquad 0 < x < 1. \quad (5.3)$$

The first function was chosen because of the similarity of its shape to that of the logistic function, which is commonly used in growth studies. The second function has a constant second derivative, and the bias of the kernel estimator is constant over $x$ in the interval $(h, 1 - h)$. By using polynomial functions, we obtained exact expressions for $C(h)$ given by (5.1) for the two functions.

*Table 1. Optimal Bandwidth and Relative Efficiency for the Kernel Estimator of $f(x) = 10x^3 - 15x^4 + 6x^5$ at $x = .789$ From a Continuous Realization*

| | $\rho$ | | |
|---|---|---|---|
| $m$ | .1 | .5 | .9 |
| | *Optimal bandwidth* | | |
| 1 | .223 | .238 | .230 |
| 4 | .202 | .216 | .212 |
| 16 | .149 | .173 | .140 |
| 64 | .110 | .121 | .089 |
| | *Relative efficiency* | | |
| 1 | .128 | .340 | .809 |
| 4 | .158 | .381 | .843 |
| 16 | .207 | .468 | .895 |
| 64 | .269 | .567 | .931 |

*Table 2. Optimal Bandwidth and Relative Efficiency for the Kernel Estimator of $f(x) = (x - \frac{1}{2})^2$ at $x = \frac{1}{2}$ From a Continuous Realization*

| | $\rho$ | | |
|---|---|---|---|
| $m$ | .1 | .5 | .9 |
| | *Optimal bandwidth* | | |
| 1 | .226 | .274 | .260 |
| 4 | .171 | .200 | .171 |
| 16 | .128 | .144 | .111 |
| 64 | .095 | .101 | .072 |
| | *Relative efficiency* | | |
| 1 | .140 | .339 | .816 |
| 4 | .183 | .424 | .873 |
| 16 | .237 | .519 | .914 |
| 64 | .303 | .615 | .944 |

The value of $\sigma$ was set to .25 and .0625 for the two cases. These values are one quarter of the range of the respective functions as $x$ varies from 0 to 1. This represents data with a moderate amount of noise. To look at the effect of various amounts of correlation, the Ornstein–Uhlenbeck correlation function $\rho(v) = \exp(-\alpha|v|)$ was used for varying $\alpha$. The values of $\alpha$ were chosen to represent correlations of $\rho = .1, .5$, and .9 in the error process for values of $x$ separated by .05. Thus $\mathrm{corr}(\varepsilon(x), \varepsilon(x + .05)) = .1, .5$, or .9. The numbers of experimental units considered were $m = 1, 4, 16, 64$. The results for function (5.2) at the point $x = .789$ are given in Table 1. This value of $x$ was chosen because it maximizes $(f''(x))^2$ and thus maximizes the bias. The results for function (5.3) at the point $x = .5$ appear in Table 2. We note that the values in Table 2 actually hold for any $x$ in $(h, 1 - h)$, where $h$ is the optimal bandwidth.

The tables first present the optimal bandwidth for various $m$ and $\rho$. The second part of each table presents a measure of the relative efficiency of using the best unbiased estimator $\bar{y}$ to that of using the kernel estimator with the optimal bandwidth. This relative efficiency can be written as $\mathrm{RE}(h) = C(h)/(\sigma^2/m)$.

For both functions, the optimal bandwidth tends to decrease as $m$ increases. The effect of varying $\rho$ on the optimal bandwidth is not so apparent. In all cases, the largest value of the optimal bandwidth occurs at the intermediate value of $\rho$. It can be argued that as $\rho$ tends to either 0 or 1, the optimal bandwidth tends to 0.

For both functions, the relative efficiency at the optimal bandwidth increases as the sample size increases. In addition, the relative efficiency increases as $\rho$ increases. Thus there is not a major improvement because of smoothing for large $\rho$, particularly for large $m$. For small $m$ and $\rho$ at .1, however, the gain in efficiency in using the kernel estimator can be tremendous.

We also examined the behavior of the optimal bandwidth when the number, $n$, of values of the independent variable, $x_i$, is finite. In this case, the MSE at a given $x$ can be computed exactly as follows.

Letting $x_i = (i - .5)/n$ $(i = 1, \ldots, n)$ be the design points, we have $s_i = i/n$ $(i = 0, \ldots, n)$. Define

$$\mathbf{f}' = (f(x_1), \ldots, f(x_n)),$$

$$\bar{\mathbf{y}}' = (\bar{y}(x_1), \ldots, \bar{y}(x_n)),$$

and

$$\mathbf{w}' = (w_1, \ldots, w_n),$$

where

$$w_i = \int_{s_{i-1}}^{s_i} K\left(\frac{x-u}{h}\right) du \Big/ \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} K\left(\frac{x-u}{h}\right) du.$$

Then the kernel estimate can be written as $\hat{f}_h^*(x) = \mathbf{w}'\bar{\mathbf{y}}$. As noted in Section 3, this choice of weights ensures that the estimate is always a weighted average of $\bar{y}_i$'s. Let $\Sigma$ denote the covariance matrix of $(\bar{\varepsilon}(x_1), \ldots, \bar{\varepsilon}(x_n))'$. The MSE at a given $x$ for a given bandwidth $h$ may be written as

$$\mathrm{MSE}(\hat{f}_h^*(x)) = \mathbf{w}'\Sigma\mathbf{w} + (f(x) - \mathbf{w}'\mathbf{f})^2. \quad (5.4)$$

The details of this part of the study are similar to the earlier part. The Epanechnikov kernel was used, and the functions of interest were (5.2) and (5.3). Since the results for the two functions are similar to each other, we present only the results for (5.2). The number of $x_i$ values was set at $n = 20$, and the number of experimental units, $m$, was set at 1, 4, 16, and 64. The covariance matrix had typical entry $\sigma_{ij} = (.0625)m^{-1}\rho^{|i-j|}$, for $\rho = .1, .5$, and $.9$. This corresponds to the correlation structure of the earlier error process sampled at points separated by .05. The independence case with $\Sigma = (.0625)m^{-1}I$ was also considered. The results for $x = .789$ are given in Table 3.

The table first presents the optimal bandwidth for various $m$ and $\rho$. These bandwidths are similar to those in Table 1 for the continuous record case. In Table 3, there is also an optimal bandwidth given for $\rho = 0$. For a given $m$, the optimum bandwidths increase and then decrease as $\rho$ ranges from 0 to .9. It is interesting that when $m = 16$ or 64, the optimum bandwidth for $\rho = .9$ is smaller than it is for $\rho = 0$. This confirms our earlier claim that correlation sometimes calls for a smaller than usual bandwidth. We also note that the bandwidths in Table 3 agree nicely with what inequality (4.3) suggested (see the discussion in Remark 4, Sec. 4).

*Table 3. Optimal Bandwidth and Relative Efficiency for the Kernel Estimator of $f(x) = 10x^3 - 15x^4 + 6x^5$ at $x = .789$ When $n = 20$*

| | $\rho$ | | | |
|---|---|---|---|---|
| $m$ | .0 | .1 | .5 | .9 |
| | | *Optimal bandwidth* | | |
| 1 | .246 | .253 | .283 | .269 |
| 4 | .208 | .213 | .226 | .214 |
| 16 | .153 | .157 | .173 | .142 |
| 64 | .109 | .111 | .116 | .094 |
| | | *Relative efficiency* | | |
| 1 | 1 | .999 | .979 | .979 |
| 4 | 1 | .998 | .986 | .999 |
| 16 | 1 | .998 | .988 | .998 |
| 64 | 1 | 1.000 | .998 | .989 |

To see the effect of correlation on the MSE, a measure of relative efficiency is presented in Table 3. This measure is $\mathrm{MSE}(h_{\mathrm{opt}})/\mathrm{MSE}(h_{\mathrm{ind}})$, where $\mathrm{MSE}(h)$ is the MSE of the kernel estimator at the given $m$ and $\rho$. The ratio compares the MSE when using the optimal bandwidth under correlation ($h_{\mathrm{opt}}$) with the MSE obtained when using the optimal bandwidth under the assumption of independence ($h_{\mathrm{ind}}$). We note that as $n \to \infty$, $h_{\mathrm{ind}} \to 0$, and $h_{\mathrm{opt}}$ tends to the value given in Table 1. In addition, the relative efficiencies will tend to those in Table 1. The results in Table 3 point out that the loss in efficiency due to assuming independence is relatively small when the number of settings of $x$ equals 20. If we take a larger $n$, $h_{\mathrm{ind}}$ would be smaller, and $h_{\mathrm{opt}}$ would remain roughly the same size. The relative efficiency would be smaller, and one would lose more by assuming independence when the errors are actually correlated.

The final part of the numerical study was designed as a check on the reasonableness of the data analysis in Section 3. The function

$$f(x) = 93.80844 + 35.027026x - 9.58274x^2 + .91072x^3 - .02847x^4$$

was chosen for the study. The shape of the function is similar to that of the kernel estimates in Figure 2. There are peaks near 3 and 12, with a local minimum near 9.

The MASE (3.2) is given by

$$M(h) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{MSE}(\hat{f}_h^*(x_j)).$$

Each term in the sum is then evaluated using (5.4). To make the situation correspond to that in Section 3, the Epanechnikov kernel was used, and the design points were $x_i = i$ $(i = 1, \ldots, 14)$. The optimal bandwidth was determined for independent errors and correlated errors. The standard deviation was set at $\sigma = 19.0$, corresponding to the estimated standard deviation of a mean in the example. In the case of correlated errors, the $(i, j)$th entry of the covariance matrix was chosen to be $\rho^{|i-j|}$, where $\rho = .7$. The IMSL (1978) subroutine ZXMIN was used to minimize $M(h)$ in both cases.

In the case of uncorrelated errors, the optimal bandwidth was found to be $h = 2.02$, and in the case of correlated errors it was found to be 1.32. Thus we would be likely to oversmooth if we treated the errors as being uncorrelated when they were actually dependent. This analytical result demonstrates the plausibility of the bandwidths obtained in our data analysis and provides another instance in which correlation leads to a smaller optimum bandwidth.

## 6. DISCUSSION

We are currently investigating the efficacy of the data-based method proposed in Section 3 for choosing a bandwidth. Such investigations have been done in the uncorrelated errors setting by Härdle and Marron (1985) and Härdle, Hall, and Marron (1985).

The method we used for estimating the correlation function $\rho$ requires multiple observations at each $x$. Estimating $\rho$ is much more problematic when only one response is available at each $x$. Diggle and Hutchinson (1985) proposed methodology for this case when a spline is used to estimate the unknown function. Bates (1985) suggested that an iterative method such as that of Cochrane and Orcutt (1949) could be used to estimate autocorrelation and arrive at a smooth function estimate simultaneously. We are now investigating such a procedure in the kernel setting. It is important to point out that the situation with one response at each $x$ and correlated errors is philosophically much different from the same setting with uncorrelated errors. This is because a consistent estimate of $f$ is not possible with correlated errors. Apparently, then, the best one could hope for would be a data-based estimate that with high probability is close to, say, a minimum integrated squared error estimate.

## APPENDIX: PROOFS OF THEOREMS 1 AND 3

*Proof of Theorem 1.* We have

$$\operatorname{var}(\hat f_h(x)) = \sigma^2(mh^2)^{-1}\sum_{i=1}^n\sum_{j=1}^n\int_{s_{i-1}}^{s_i}\int_{s_{j-1}}^{s_j}\rho(x_i - x_j) \times K\left(\frac{x-u}{h}\right)K\left(\frac{x-v}{h}\right)du\,dv$$

and for $h$ sufficiently small,

$$\begin{aligned}\int_{-1}^1\int_{-1}^1 \rho(h(u-v))K(u)K(v)\,du\,dv &= h^{-2}\int_0^1\int_0^1 K\left(\frac{x-u}{h}\right)K\left(\frac{x-v}{h}\right)\rho(v-u)\,du\,dv\\ &= h^{-2}\sum_{i=1}^n\sum_{j=1}^n\int_{s_{i-1}}^{s_i}\int_{s_{j-1}}^{s_j}K\left(\frac{x-u}{h}\right)K\left(\frac{x-v}{h}\right)\rho(v-u)\,du\,dv.\end{aligned}$$

Therefore,

$$\begin{aligned}&\left|\operatorname{var}(\hat f_h(x)) - \sigma^2 m^{-1}\int_{-1}^1\int_{-1}^1\rho(h(u-v))K(u)K(v)\,du\,dv\right|\\ &\le \sigma^2(mh^2)^{-1}\sum_{i=1}^n\sum_{j=1}^n\int_{s_{i-1}}^{s_i}\int_{s_{j-1}}^{s_j}|\rho(x_i-x_j)-\rho(v-u)| \times K\left(\frac{x-u}{h}\right)K\left(\frac{x-v}{h}\right)du\,dv.\end{aligned}$$

By the Lipschitz continuity of $\rho$,

$$|\rho(x_i - x_j) - \rho(v-u)| \le B|(x_i - x_j) - (v-u)| \le B[|x_i - v| + |x_j - u|].$$

Since $s_{j-1} \le u \le s_j$ and $s_{i-1} \le v \le s_i$, we have

$$|x_i - v| + |x_j - u| \le (\tfrac{1}{2})[(s_i - s_{i-1}) + (s_j - s_{j-1})] \le \sup_k (s_k - s_{k-1}).$$

It follows that

$$\begin{aligned}&\left|\operatorname{var}(\hat f_h(x)) - \sigma^2 m^{-1}\int_{-1}^1\int_{-1}^1\rho(h(u-v))K(u)K(v)\,du\,dv\right|\\ &\le B\sigma^2(mh^2)^{-1}\sup_k(s_k - s_{k-1}) \times \sum_{i=1}^n\sum_{j=1}^n\int_{s_{i-1}}^{s_i}\int_{s_{j-1}}^{s_j}K\left(\frac{x-u}{h}\right)K\left(\frac{x-v}{h}\right)du\,dv\\ &= [B\sigma^2\sup_k(s_k - s_{k-1})/m]\left[h^{-1}\int_0^1 K\left(\frac{x-u}{h}\right)du\right]^2\\ &= 0(1/nm).\end{aligned}$$

*Proof of Theorem 3.* From our Theorem 1, equation (6) of Gasser and Müller (1984), and a standard Taylor series argument,

$$\operatorname{MSE}(\hat f_h(x)) = (\sigma^2/m)\int_{-1}^1\int_{-1}^1\rho(h(u-v))K(u)K(v)\,du\,dv + h^4\sigma_K^4(f''(x))^2/4 + 0(1/nm + h^2/n + 1/n^2) + o(h^4).$$

Since $m/n = 0(1)$, the first of the two remainder terms is actually $0((nm)^{-1} + h^2/n)$. Consider now

$$\begin{aligned}&\int_{-1}^1\int_{-1}^1[\rho(h(u-v)) - 1]K(u)K(v)\,du\,dv\\ &= 2h\int_{-1}^1\int_v^1(u-v)\left[\frac{\rho(h(u-v))-1}{h(u-v)}\right]K(u)K(v)\,du\,dv.\end{aligned}$$

Since $\rho$ is Lipschitz continuous, $|(\rho(y)-1)/y|$ is bounded; hence, by dominated convergence, the above double integral tends to $\rho'(0+)C_K/2$. Therefore,

$$\operatorname{MSE}(\hat f_h(x)) = (\sigma^2/m)(1 + \rho'(0+)C_K h) + h^4\sigma_K^4(f''(x))^2/4 + 0(1/nm + h^2/n) + o(h^4 + h/m).$$

From this expression and using the condition $m/n = 0(1)$ it follows that

$$\begin{aligned}\operatorname{MSE}(\hat f_h(x)) &\sim (\sigma^2/m)(1 + \rho'(0+)C_K h) + h^4\sigma_K^4(f''(x))^2/4\\ &= \sigma^2/m + E(m,h).\end{aligned}$$

To prove the rest of the theorem, we first point out that $E(m, h_m) < E(m,h)$ whenever $h > 0$ and $h \ne h_m$. Now,

$$\begin{aligned}E(m,h_m) &= m^{-4/3}(-\tfrac{3}{4})(\sigma^2\rho'(0-)C_K)^{4/3}/(\sigma_K^4(f''(x))^2)^{1/3}\\ &= Am^{-4/3}.\end{aligned}$$

Note that, by assumption, $A < 0$. Letting $h_{n,m}$ be as defined in Theorem 3, we have

$\operatorname{MSE}(\hat f_{h_m}(x)) < \operatorname{MSE}(\hat f_{h_{n,m}}(x))$ iff

$$A + 0\left(\frac{m^{1/3}}{n} + \frac{m^{2/3}}{n}\right) + o(1) < m^{4/3}E(m,h_{n,m}) + 0\left(\frac{m^{1/3}}{n} + \frac{m^{4/3}h_{n,m}^2}{n}\right) + o(m^{4/3}h_{n,m}^4 + m^{1/3}h_{n,m}). \quad (A.1)$$

By definition of $h_{n,m}$, it is easily verified that $A - m^{4/3}E(m,h_{n,m}) \le \delta < 0$ for all $n$ and $m$ sufficiently large. The remainder term on the left side of Inequality (A.1) tends to 0, since $m/n = 0(1)$. The remainder term on the right side of (A.1) tends to 0 if $mh_{n,m}^3 = 0(1)$, thus proving the optimality of $h_m$ for this case. Suppose now that $mh_{n,m}^3 \to \infty$ as $n, m \to \infty$. Inequality (A.1) is equivalent to

$$[m^{4/3}E(m,h_{n,m})/A]\left[1 + 0\left(\frac{1}{nm} + \frac{h_{n,m}^2}{n}\right)\Big/E(m,h_{n,m}) + o\left(h_{n,m}^4 + \frac{h_{n,m}}{m}\right)\Big/E(m,h_{n,m})\right] < 1 + o(1).$$

Using $m/n = 0(1)$ and $mh^3_{n,m} \to \infty$, it is easily verified that

$$1 + 0\left(\frac{1}{nm} + \frac{h^2_{n,m}}{n}\right) \Big/ E(m, h_{n,m}) + o\left(h^4_{n,m} + \frac{h_{n,m}}{m}\right) \Big/ E(m, h_{n,m})$$

has limit 1 as $n, m \to \infty$. In addition, by definition of $h_{n,m}$, we have $m^{4/3}E(m, h_{n,m})/A \leq 1 - \varepsilon$ $(\varepsilon > 0)$ for all $n, m$ sufficiently large, and the result is proved.

[*Received October 1985. Revised May 1986.*]

## REFERENCES

Andersen, A. H., Jensen, E. B., and Schou, G. (1981), "Two-Way Analysis of Variance With Autocorrelated Errors," *International Statistical Review*, 49, 153–157.

Azzalini, A. (1984), "Estimation and Hypothesis Testing for Collections of Autoregressive Time Series," *Biometrika*, 71, 85–90.

Bates, P. J. (1985), Discussion of "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," by B. W. Silverman, *Journal of the Royal Statistical Society*, Ser. B, 47, 32.

Cochrane, D., and Orcutt, G. H. (1949), "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms," *Journal of the American Statistical Association*, 44, 32–61.

Diggle, P. J. (1985), Discussion of "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting," by B. W. Silverman, *Journal of the Royal Statistical Society*, Ser. B, 47, 28–29.

Diggle, P. J., and Hutchinson, M. F. (1985), "Spline Smoothing With Autocorrelated Errors," CSIRO Technical Report.

Gasser, Th., and Müller, H. G. (1979), "Kernel Estimation of Regression Functions," in *Smoothing Techniques for Curve Estimation*, eds. Th. Gasser and M. Rosenblatt, Heidelberg: Springer-Verlag, pp. 23–68.

——— (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," *Scandinavian Journal of Statistics*, 11, 171–185.

Gasser, Th., Müller, H. G., Kohler, W., Molinari, L., and Prader, A. (1984), "Nonparametric Regression Analysis of Growth Curves," *The Annals of Statistics*, 12, 210–229.

Geisser, S. (1980), "Growth Curve Analysis," in *Handbook of Statistics 1*, ed. P. R. Krishnaiah, Amsterdam: North-Holland, pp. 89–115.

Ghosh, M., Grizzle, J. E., and Sen, P. K. (1973), "Nonparametric Methods in Longitudinal Studies," *Journal of the American Statistical Association*, 68, 29–36.

Grizzle, J. E., and Allen, D. M. (1969), "Analysis of Growth and Dose Response Curves," *Biometrics*, 25, 357–381.

Härdle, W., Hall, P., and Marron, J. S. (1985), "How Far Are Automatically Chosen Regression Smoothing Parameters From Their Optimum?," Mimeo Series 1589, University of North Carolina, Dept. of Statistics.

Härdle, W., and Marron, J. S. (1985), "Optimal Bandwidth Selection in Nonparametric Regression Function Estimation," *The Annals of Statistics*, 13, 1465–1481.

Härdle, W., and Tuan, P.-D. (1986), "Some Theory on *M*-Smoothing of Time Series," *Journal of Time Series Analysis*, 7, 191–204.

International Mathematical and Statistical Libraries (1978), *IMSL Manual* (Vols. 1 and 2), Houston, TX: Author.

Masry, E. (1983), "Spectral and Probability Density Estimation From Irregularly Observed Data," in *Time Series Analysis of Irregularly Observed Data*, ed. E. Parzen, Heidelberg: Springer-Verlag, pp. 224–250.

Müller, H.-G. (1984), "Optimal Designs for Nonparametric Kernel Regression," *Statistics and Probability Letters*, 2, 285–290.

Potthoff, R. F., and Roy, S. N. (1964), "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curve Problems," *Biometrika*, 51, 313–326.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation*, Orlando, FL: Academic Press.

Rao, C. R. (1965), "The Theory of Least Squares When the Parameters Are Stochastic and Its Application to the Analysis of Growth Curves," *Biometrika*, 52, 447–458.

Reinsel, G. (1982), "Multivariate Repeated-Measurement or Growth Curve Models With Multivariate Random-Effects Covariance Structure," *Journal of the American Statistical Association*, 77, 190–195.

Rice, J. (1984a), "Bandwidth Choice for Nonparametric Regression," *The Annals of Statistics*, 12, 1215–1230.

——— (1984b), "Boundary Modification for Kernel Regression," *Communications in Statistics—Theory and Methods*, 13, 893–900.